OSCARS
Open Science Clusters' Action for Research & Society

Funded Project

Fail2Fair: Recovering Discarded Macromolecular Crystallographic Data

Implemented by

ibmb | CSIC | PSI | Universität Konstanz

Funded by the European Union

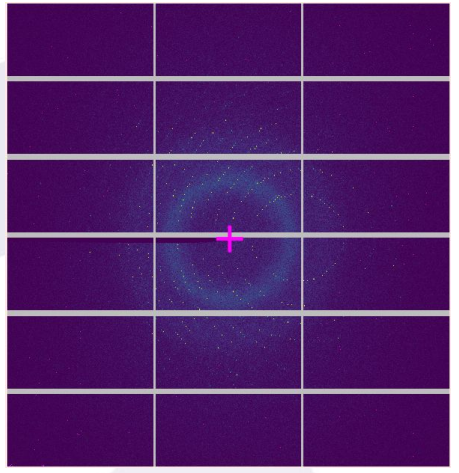
CHALLENGE ADDRESSED

- Macromolecular crystallography (MX) provides vital data for structure biology with applications in health and biotechnology
- Failed MX experiments are rarely archived
- Missing failed datasets limit method development and AI training
- Exponential increase in data rates for MX
- Archiving all X-ray images prohibitively expensive

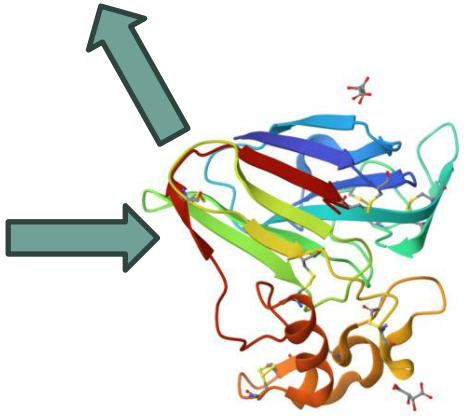
RCSB PDB PROTEIN DATA BANK

250,079 Structures from the PDB archive

1,068,577 Computed Structure Models (CSM)



X-ray diffraction images

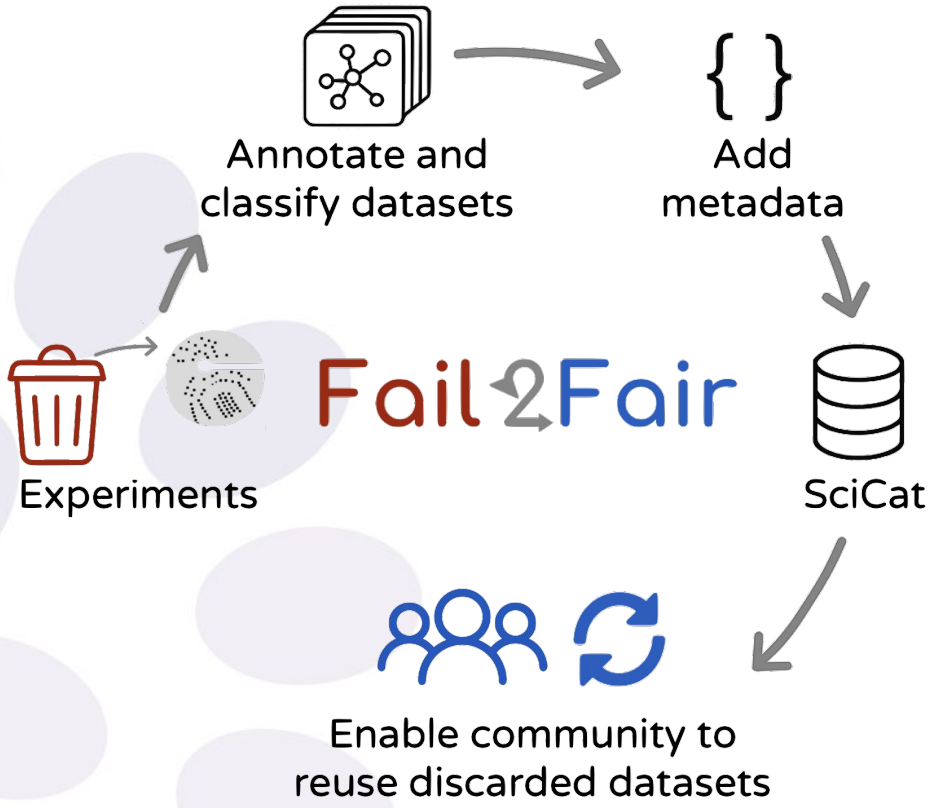


Protein structure

Synchrotron facility

SOLUTION

- Extend metadata models and automate dataset classification
- Compression to allow sustainable archiving and reuse of all datasets
- Engage community feedback and involvement with two workshops

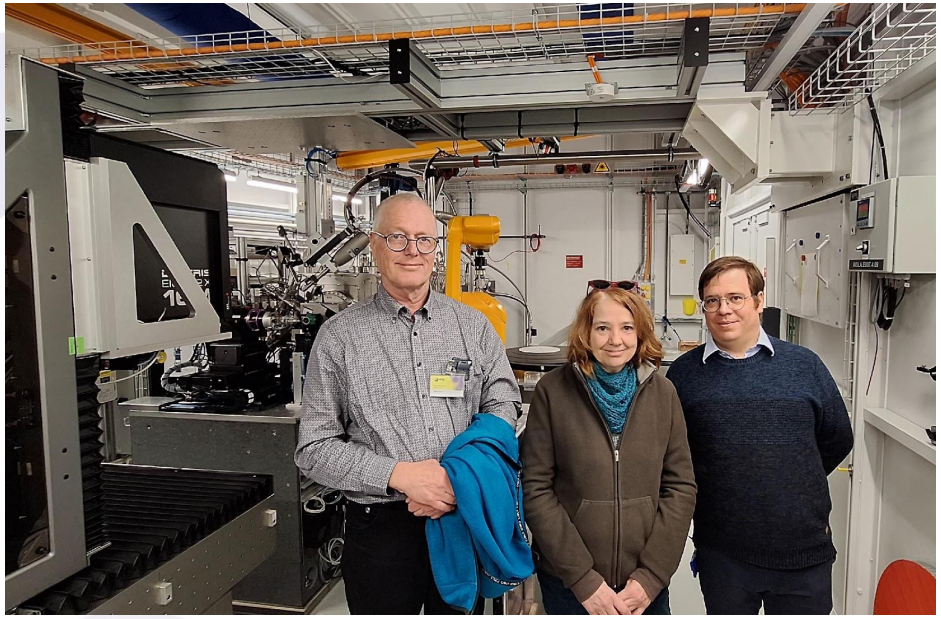


RESULTS

- We will improve FAIR data availability at synchrotron facilities
- We will provide software to annotate datasets for common pathologies
- We will provide a proof of principle for scientific applications using these data

TEAM

Institute of Molecular Biology of Barcelona (ES)
Isabel Usón
Iracema Caballero
Josep Triviño

Paul Scherrer Institute (CH)
Filip Leonarski

University of Konstanz (DE)
Kay Diederichs